



I should remind you that it is EMBO Journal policy to allow a single round of revision only and that, therefore, acceptance or rejection of the manuscript will depend on the completeness of your responses included in the next, final version of the manuscript. When you submit a revised version to the EMBO Journal, please make sure you upload a letter of response to the referees' comments. Please note that when preparing your letter of response to the referees' comments that this will form part of the Review Process File, and will therefore be available online to the community. For more details on our Transparent Editorial Process initiative, please visit our website: <http://www.nature.com/emboj/about/process.html>

Thank you for the opportunity to consider your work for publication. I look forward to your revision.

Yours sincerely,

Editor  
The EMBO Journal

-----  
REFEREE COMMENTS

Referee #1

In this study the authors analyzed the histone H3 methylation state during different stages of T cell development. Specifically, they systematically analyzed the three different methylation states of histone H3K4 to determine if H3K4me1-marked regions, that are commonly believed to represent enhancers, acquire H3K4me2/3 upon developmental transitions.

To this aim, they analyzed freshly purified thymocytes from Rag2<sup>-/-</sup> mice before and after injection of an anti-TCR antibody, which stimulates progression of thymocytes from the double-negative to the double-positive stage. These cells were used for ChIP-chip experiments using antibodies that recognize H3K4me1/2/3 or other histone modifications that correlate with either gene activation or repression. The array for hybridization was designed to contain genomic regions (50 kb each) corresponding to genes that are differentially expressed during T cell development.

The main finding of the study was the detection of an increase in H3K4me2 or me3 at enhancers contained in genomic regions that underwent activation during progression from DN to DP. Similarly, enhancers in genes that were repressed during differentiation lost H3K4me3. The distal regions associated with H3K4me3 were generally associated with RNA polymerase II and in the few cases tested intergenic non-coding RNAs were detected. Moreover, RNA polymerase II stalling induced by inhibition of Cdk9 using specific inhibitors resulted in a proximal increase in the H3K4me3 signal.

Based on these data the authors argue that H3K4me2/3 discriminates between active and poised (but inactive) enhancers.

The authors should be commended for attempting at dissecting the enhancer repertoire (which is a pertinent and actual issue in the field) and for doing it in a relevant primary cell system. However, there are some technical and conceptual issues with the study that in my opinion make some of the conclusions of the paper difficult to accept.

First, nearly all data have been generated using array-based hybridization approaches. As compared to high-throughput sequencing, arrays have a much lower dynamic range and quantitative differences may remain unappreciated or blunted. This is a relevant issue here because it is unclear whether the H3K4me3 signal seen at enhancers is comparable to the one detected at promoters. In Fig. 3, Q-PCR data indicate a difference of about one order of magnitude between H3K4me1 and H3K4me3 at a T cell enhancer, but when one looks at all the other regions in Fig 1 and 2, the K4me1 signal appears in general lower than or similar to the H3K4me3. Several published data sets showed that enhancers may be associated with a low-level enrichment of H3K4me3, but using a quantitative assay this enrichment appears to be much lower than that detected at active promoters. Another compelling technical issue is that in some cases the H3K4me3 signal appears to be spread way beyond what is usually observed: for instance the K4me3-positive region in Fig 1A spans more

than 20kB on the Cd3d and Cd3g genes and even more than that on Cd3e; in Fig 1C there is a long K4me3-positive region of almost 100kB at the 5' of the gene and additional very long regions inside the gene, including the one labeled as enhancer); in Fig 2A the H3K4me3 region extends almost without interruption along the whole locus. However, the common finding in most studies is that H3K4me3 is restricted to a few nucleosomes after the transcription start sites, and this makes me think that there are some important technical issues (either with sample sonication or hybridization) that negatively impact on the resolution of the experiments shown here.

Taking these two issues together, I am afraid that these data may not be considered robust enough by the scientific community, and therefore they should be strengthened and improved to eliminate any reasonable technical concern. The simplest way to do so would be to show Q-PCR data for a substantial number of the genomic regions in the array. Alternatively, ChIP-Seq would be the optimal and definitive solution.

The second major issue I do find, is more conceptual but extremely relevant. The issue is the following: what do we define an "enhancer" and what do we define a "gene"? Nowadays, after the ENCODE project, any region located outside protein-coding genes, associated with H3K4me3 and that generates transcripts, fits pretty well the definition of a non-coding RNA gene. Therefore, a most reasonable interpretation of these data is that during T cell differentiation there are some ncRNA genes that are either activated or repressed, leading to corresponding changes in H3K4me3 levels within H3K4me1-marked regions (noting that also many inactive genes that are not permanently silenced do contain H3K4me1 before being activated). So, it may well be that the authors are not distinguishing active from inactive enhancers using H3K4me2/3, but they are simply identifying which ones of the H3K4me1-associated regions are enhancers according to the classic definition, and which ones are conversely non-coding RNA genes.

## Referee #2

In eukaryotes, combinations of post-translational histone modifications and transcription factor occupancy determine the chromatin landscape during cell development and differentiation. Several epigenomic studies have identified a set of histone modifications that preferentially associate with transcriptional regulatory regions, including enhancers. In general, it was accepted that distinct stoichiometries of H3K4 methylation preferentially mark active promoters (H3K4me3) and intergenic enhancers (H3K4me1/2). Little is known about how these modifications regulate the function and activation of these regulatory elements.

In this study, Pekowska et al. examine the pattern of histone marks at loci containing genes that are specifically expressed in developing lymphocytes, with a focus on early thymocyte development. The authors conclude that, in addition to H3K4me1, the di- and tri-methyl marks are present at tissue-specific enhancers, revealing a better chromatin signature with which to identify new regulatory elements. In addition, the authors find that Pol II is bound to nearly all of the examined enhancers and may be responsible (indirectly) for deposition of the H3K4me3 mark.

In general, the experimental data are informative and, in most cases, include important controls. However, the study remains fairly descriptive and the major conclusion - that H3K4me1/2/3 constitutes a more useful chromatin signature for enhancers - is not convincing.

## Specific comments:

1. Although I agree that the H3K4me1/2/3 + Pol II signature seems to be present at most enhancers, at least one of the enhancers that the authors highlight ( $E_{\mu}$ ) is known to contain a strong promoter that should be enriched for H3K4me3. With regard to the utility of this signature for identification of new elements, many non-enhancer regions bear the same pattern of H3K4 modifications in the loci presented (to name a few, Fig. 1C throughout the *Ikzf1* locus, additional peaks 5' to *Dntt*, Fig. 2A many between *Cd8b1* and *Cd8a*). Without the boxes around known enhancers, I would have been hard pressed to pick out the regulatory element in many loci shown. Have the authors employed this approach to identify several new enhancers? Candidates are highlighted but not tested.

2. A related concern is that the authors used crosslinked chromatin for their ChIP-chip studies. Many enhancers become crosslinked to interacting promoters using this approach. Therefore, it remains unclear whether each of the marks (especially H3K4me3) is on enhancer-associated nucleosomes or

on the promoter nucleosomes that get crosslinked to the enhancer. If the authors wish to convincingly demonstrate that H3K4me2/3 are enhancer-intrinsic marks, they must perform at least some ChIP-chip studies using mononucleosome preparations from native chromatin.

3. There is very little difference in the patterns observed for the H3K4me2 versus me3 modification. Were there any controls for antibody specificity?

4. The Pol II studies in Fig. 5 are rather superficial. Studies with antibodies that recognize phosphorylated forms of the polymerase (Ser2 versus Ser5) would be more informative. In addition, the authors should include data for an elongation mark (H3K36me) in their Pol II inhibitor experiments.

5. There are several Figures where controls for tissue specificity would be useful. In Fig. 3B, the scales for H3K4me1 and me3 data are very different. How do these data look when compared with a non-thymocyte (or non-lymphocyte)? Similarly, the modification patterns on Cd3d look quite similar in B versus T lineage cells. Non-lymphocytes apparently were not examined.

6. A subset of data should be supported by focused ChIP-PCR assays. Most relevant are those shown in Fig. 4, where differences are especially subtle. Are these differences reproducible when performed in a more quantitative manner and how do these regions compare with marks in non-lymphoid cells?

#### Referee #3

In this study, Pekowska et al. nicely document a correlation between histone H3K4 dimethylation and trimethylation and activity of stage-specific enhancers. Figure 1B is particularly revealing, by showing that the previously proposed selective enrichment of H3K4me1 at enhancers is lost when examining stage-specific enhancers. This figure was derived from a genome-wide compilation of modification levels, but the authors then validate this finding by examining individual loci using high-density chip-on-chip tiling arrays. The data provide strong evidence that the authors conclusions are valid at at least a subset of enhancers. The authors also provide a possible connection between H3K4me3 modification of enhancers and polymerase association.

This manuscript is highly interesting and will represent an important addition to the literature. It is particularly important because previously published data have led to the erroneous view that all enhancers contain high levels of H3K4me1, but low levels of H3K4me3. However, as with most genome-wide studies, the published studies described statistical averages and trends. By separating stage-specific enhancers from enhancers that are more constitutively active, the authors of the current manuscript show that the H3K4me1/H3K4me3 ratio is much lower at active enhancers. Since many labs are currently identifying enhancers on the basis of a high ratio, this paper will of considerable value.

My only suggestion is for the authors to discuss more extensively the observation that H3K4me3 appears to extend through a surprisingly broad region of the active loci. Has this been observed by others? Doesn't the broad distribution raise questions about the hypothesized close link between H3K4me3 and polymerase recruitment?

**Point-by-point responses to referees****Referee #1**

*In this study the authors analyzed the histone H3 methylation state during different stages of T cell development. Specifically, they systematically analyzed the three different methylation states of histone H3K4 to determine if H3K4me1-marked regions, that are commonly believed to represent enhancers, acquire H3K4me2/3 upon developmental transitions.*

*To this aim, they analyzed freshly purified thymocytes from Rag2<sup>-/-</sup> mice before and after injection of an anti-TCR antibody, which stimulates progression of thymocytes from the double-negative to the double-positive stage. These cells were used for ChIP-chip experiments using antibodies that recognize H3K4me1/2/3 or other histone modifications that correlate with either gene activation or repression. The array for hybridization was designed to contain genomic regions (50 kb each) corresponding to genes that are differentially expressed during T cell development.*

*The main finding of the study was the detection of an increase in H3K4me2 or me3 at enhancers contained in genomic regions that underwent activation during progression from DN to DP. Similarly, enhancers in genes that were repressed during differentiation lost H3K4me3. The distal regions associated with H3K4me3 were generally associated with RNA polymerase II and in the few cases tested intergenic non-coding RNAs were detected. Moreover, RNA polymerase II stalling induced by inhibition of Cdk9 using specific inhibitors resulted in a proximal increase in the H3K4me3 signal.*

*Based on these data the authors argue that H3K4me2/3 discriminates between active and poised (but inactive) enhancers.*

*The authors should be commended for attempting at dissecting the enhancer repertoire (which is a pertinent and actual issue in the field) and for doing it in a relevant primary cell system. However, there are some technical and conceptual issues with the study that in my opinion make some of the conclusions of the paper difficult to accept.*

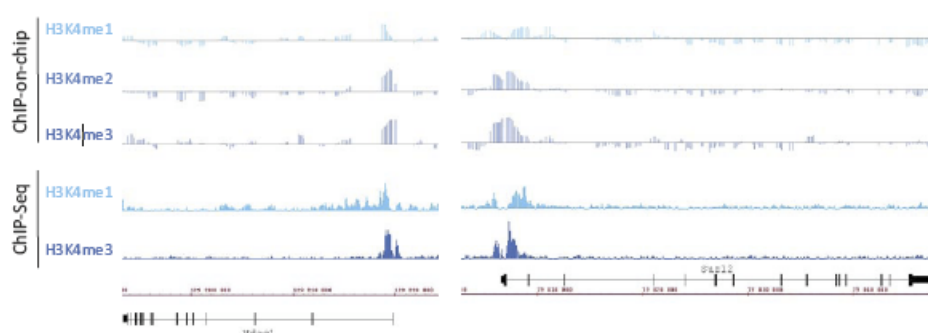
*First, nearly all data have been generated using array-based hybridization approaches. As compared to high-throughput sequencing, arrays have a much lower dynamic range and quantitative differences may remain unappreciated or blunted. This is a relevant issue here because it is unclear whether the H3K4me3 signal seen at enhancers is comparable to the one detected at promoters. In **Fig. 3**, Q-PCR data indicate a difference of about one order of magnitude between H3K4me1 and H3K4me3 at a T cell enhancer, but when one looks at all the other regions in **Fig 1** and **2**, the K4me1 signal appears in general lower than or similar to the H3K4me3. Several published data sets showed that enhancers may be associated with a low-level enrichment of H3K4me3, but using a quantitative assay this enrichment appears to be much lower than that detected at active promoters.*

**Response 1:** We are aware of the technical limits of array-based approaches and have taken into account the referee's comments to strengthen our data (see **Responses 2-3** below). Concerning the **Figure 3**, we do apologize for the confusion as we realized that we previously made a computation mistake when quantifying the enrichment levels by qPCR. We now provide the corrected values in a new figure 3B. We also provide additional controls, including ChIP experiments in a non-lymphoid tissue (ES cells; also suggested by Referee #2) and qPCR analyses of an active promoter (*Cd3e*). We stress the fact that although H3k4me3 enrichment observed at the *Tcrb* enhancer (Eb) in DRag thymocytes is relatively low as compared to the *Cd3e* promoter, this enrichment is significant compared to DRagEb169 thymocytes and ES cells. We hope the reviewer #1 will be satisfied by the new Figure 3.

*Another compelling technical issue is that in some cases the H3K4me3 signal appears to be spread way beyond what is usually observed: for instance the K4me3-positive region in **Fig 1A** spans more than 20kb on the *Cd3d* and *Cd3g* genes and even more than that on *Cd3e*; in **Fig 1C** there is a long*

*K4me3-positive region of almost 100kB at the 5' of the gene and additional very long regions inside the gene, including the one labeled as enhancer); in Fig 2A the H3K4me3 region extends almost without interruption along the whole locus. However, the common finding in most studies is that H3K4me3 is restricted to a few nucleosomes after the transcription start sites, and this makes me think that there are some important technical issues (either with sample sonication or hybridization) that negatively impact on the resolution of the experiments shown here.*

**Response 2:** In a previous publication we have made the observation that a subset of tissue specific genes harbor a broad distribution of H3K4 methylation, including H3K4me3 (Genome Res (2010) 20, 1493-502). This observation was made based on published ChIP-Seq data set from human CD4 T cells and mouse Brain and is in full agreement with our current data. We stress the fact that the broad profile is not observed at all expressed genes as illustrated by two examples of ubiquitously expressed genes (**additional Figure 1**). However, to completely rule out the possibility of a technical artifact of our ChIP-on-chip experiments, we performed H3K4me1 and me3 ChIP-Seq experiments using uncrosslinked MNase-digested samples (also suggested by Referee #2). As shown in the new Supplementary figures S4, our ChIP-Seq data corroborate the observation of broad distribution of methylated H3K4 at tissue specific loci (see also additional Figure 1 for examples of ChIP-Seq profiles at ubiquitously expressed genes). These findings are mentioned in the results section (p. 6, lines 17-19) and discussed in p. 18, lines 1-15.



**Additional Figure 1:** Examples of histone methylation profiles at two ubiquitously-expressed genes.

*Taking these two issues together, I am afraid that these data may not be considered robust enough by the scientific community, and therefore they should be strengthened and improved to eliminate any reasonable technical concern. The simplest way to do so would be to show Q-PCR data for a substantial number of the genomic regions in the array. Alternatively, ChIP-Seq would be the optimal and definitive solution.*

**Response 3:** As mentioned, we have confirmed our ChIP-on-chip results by using ChIP-Seq (new supplementary **Fig. S4 and S9**). Moreover, in a few cases, we also performed ChIP-qPCR analyses to validate enrichment differences observed by ChIP-on-chip (new supplementary **Fig. S3 and S8**; see also **Response 6** to Referee #2). Overall, these results truly support our main conclusions. We hope Referee #1 is satisfied by the new set of data.

*The second major issue I do find, is more conceptual but extremely relevant. The issue is the following: what do we define an "enhancer" and what do we define a "gene"? Nowadays, after the ENCODE project, any region located outside protein-coding genes, associated with H3K4me3 and that generates transcripts, fits pretty well the definition of a non-coding RNA gene. Therefore, a most reasonable interpretation of these data is that during T cell differentiation there are some ncRNA genes that are either activated or repressed, leading to corresponding changes in H3K4me3 levels within H3K4me1-marked regions (noting that also many inactive genes that are not permanently silenced do contain H3K4me1 before being activated). So, it may well be that the authors are not distinguishing active from inactive enhancers using H3K4me2/3, but they are simply identifying which ones of the H3K4me1-associated regions are enhancers according to the classic definition, and which ones are conversely non-coding RNA genes.*

**Response 4:** We agree with the referee that this is an important issue which is currently the focus of extensive research. Indeed, the definition of enhancers and their functions is a rapidly evolving field (see the recent reviews on the topic: Ong and Corces (2011) *Nat. Rev. Genet.* 12; 283-293; Bulger and Groudine (2011) *Cell* 144; 327-339), and illustrated, for instance, by the recent finding of enhancer-associated non-coding RNAs (i.e. Nature (2010) 465(7295):182-7; *PLoS Biol.* (2010) 8(5):e1000384; this study). Thus, to specifically address the conceptual issue highlighted by Referee #1 concerning the difficulty of defining intergenic H3K4me3 enriched regions as enhancers, we provide now two sets of analyses/results. On the one hand, we have analyzed the relative enrichment of H3K4me1 and H3K4me3 at genomic regions with overlapping H3K4me1 and H3K4me3 peaks in function of their genomic locations (new Supplementary Fig. S7; p. 11, lines 11-14). These analyses show that, even in the presence of H3K4me3, H3K4me1 (which is a hallmark of enhancers) is higher at intergenic regions compared to TSS-associated regions. On the other hand, we have performed luciferase reporter assays to test for enhancer activity of several genomic regions associated with both H3K4me1 and H3K4me3 peaks or with H3K4me1 peaks alone. As shown in the new Supplementary Fig. S9, only genomic regions associated with H3K4me1/3 peaks displayed enhancer activity in our assay (p. 11, line 11 to p. 12, line 3). Overall, we think that, together with these new set of data/analyses, our manuscript has extensively addressed this issue. This is discussed in a dedicated paragraph in the discussion section (p. 17, lines 5-12):

*“Firstly, the activity of all the genuine T-cell enhancers studied here was associated with the presence of H3K4me2/3 (Figures 2, S4 and S5). Secondly, the level of H4K4me1, which is expected to be enriched at enhancers, was found to be higher at intergenic H3K4me1/me3 domains as compared to TSS-associated domains (Figure S7). Thirdly, several intergenic H3K4me1/me3 regions demonstrated significant enhancer activity in a luciferase reporter assays (Figure S9). Finally, a strong correlation was found to exist between distal H3K4me1 domains that either acquired or lost H3K4me2/3 and, respectively, the induction or repression of neighboring genes by pre-TCR signalling (Figure 5B).”*

## Referee #2

*In eukaryotes, combinations of post-translational histone modifications and transcription factor occupancy determine the chromatin landscape during cell development and differentiation. Several epigenomic studies have identified a set of histone modifications that preferentially associate with transcriptional regulatory regions, including enhancers. In general, it was accepted that distinct stoichiometries of H3K4 methylation preferentially mark active promoters (H3K4me3) and intergenic enhancers (H3K4me1/2). Little is known about how these modifications regulate the function and activation of these regulatory elements.*

*In this study, Pekowska et al. examine the pattern of histone marks at loci containing genes that are specifically expressed in developing lymphocytes, with a focus on early thymocyte development. The authors conclude that, in addition to H3K4me1, the di- and tri-methyl marks are present at tissue-specific enhancers, revealing a better chromatin signature with which to identify new regulatory elements. In addition, the authors find that Pol II is bound to nearly all of the examined enhancers and may be responsible (indirectly) for deposition of the H3K4me3 mark.*

*In general, the experimental data are informative and, in most cases, include important controls. However, the study remains fairly descriptive and the major conclusion - that H3K4me1/2/3 constitutes a more useful chromatin signature for enhancers - is not convincing.*

## *Specific comments:*

*1. Although I agree that the H3K4me1/2/3 + Pol II signature seems to be present at most enhancers, at least one of the enhancers that the authors highlight (Em) is known to contain a strong promoter that should be enriched for H3K4me3. With regard to the utility of this signature for identification of new elements, many non-enhancer regions bear the same pattern of H3K4 modifications in the loci presented (to name a few, Fig. 1C throughout the Ikzf1 locus, additional peaks 5' to Dntt, Fig. 2A many between Cd8b1 and Cd8a). Without the boxes around known enhancers, I would have been hard pressed to pick out the regulatory element in many loci shown. Have the authors*

*employed this approach to identify several new enhancers? Candidates are highlighted but not tested.*

**Response 1:** To avoid any ambiguity concerning the examples of known enhancers, we have removed Em from **Fig. S5A**. Concerning the second remark, as discussed in our manuscript (**p. 18, lines 1-15**), highly tissue specific loci appear to be strongly enriched in H3K4 methylations with enrichment spreading out the known enhancers. We stress the fact that these genes often harbor very complex regulatory regions with many DNase I hypersensitive sites (DHS). For instance, in the mouse, 10 DHS have been described throughout the gene body of the *Ikzf1* gene (EMBO J. (2003) 22(9): 2211-23) and 10 around the *Dnrt* locus (Mol. Immunol. (2008) 45:1009-17). For simplicity, we have shown in the corresponding figures only the DHS matching to well characterized enhancers. It is therefore possible that other less well characterized regions might harbor cis-regulatory functions. To further investigate whether the presence of H3K4me3 might help to identified new enhancers, we performed luciferase reporter assays to test for potential enhancer activities at intergenic regions associated with H3K4me1 peaks overlapping or not H3K4me3 peaks. As shown in the new **Fig. S9** and described in the result section (**p. 11, line 11 to p. 12, line 3**), most of the regions associated with both H3K4me1 and H3K4me3 peaks displayed enhancer activity in a T cell line, whereas regions associated with only H3K4me1 did not.

*2. A related concern is that the authors used crosslinked chromatin for their ChIP-chip studies. Many enhancers become crosslinked to interacting promoters using this approach. Therefore, it remains unclear whether each of the marks (especially H3K4me3) is on enhancer-associated nucleosomes or on the promoter nucleosomes that get crosslinked to the enhancer. If the authors wish to convincingly demonstrate that H3K4me2/3 are enhancer-intrinsic marks, they must perform at least some ChIP-chip studies using mononucleosome preparations from native chromatin.*

**Response 2:** To address this sound request we performed ChIP-Seq experiments using mononucleosome preparations from native chromatin. Strikingly, we still observed H3K4me3 enrichment at known enhancers which are active in thymocytes (new supplementary **Fig. S4**), demonstrating that H3K4me3 is truly present on enhancer-associated nucleosomes. We thank the Referee #2 for suggesting this experiment which, we believe, has strengthened the main message of the manuscript. This new set of data is now described in the result section (**p.9, lines 10-14**).

*3. There is very little difference in the patterns observed for the H3K4me2 versus me3 modification. Were there any controls for antibody specificity?*

**Response 3:** Although we agree with Referee #2 that H3K4me2 and me3 display related patterns, this may not be so surprising as both marks have been linked to gene activation (i.e. Cell (2007) 129, 823-37). However, it is of note that this is not always the case. In our study these two marks displayed different profiles around the TSS (**Fig. S2C**) and there are many examples where the patterns observed for H3K4me2 and H3K4me3 are different (see the examples in **Fig. 4C**). More generally we have found that intergenic regions generally gain only one methylation mark at a time (see **Fig. S6**), supporting the fact that the enrichment profile obtained with each antibody is specific. Lastly, these two antibodies have been used in other comparative studies with similar results (i.e. Cell (2007) 129, 823-37).

*4. The Pol II studies in Fig. 5 are rather superficial. Studies with antibodies that recognize phosphorylated forms of the polymerase (Ser2 versus Ser5) would be more informative. In addition, the authors should include data for an elongation mark (H3K36me) in their Pol II inhibitor experiments.*

**Response 4:** To take into account the referee's comments, we have performed additional genome-wide analyses comparing H3K4me3 and Pol II enrichments using the ChIP-Seq data obtained from DRagCD3 thymocytes. As shown in the new **Figures 5E and S11**, the presence of Pol II at H3K4me1 peaks is strictly dependent on H3K4me3 (**p. 13, lines 6-8**). While we agree that potentially interesting findings could emerge in studying the phosphorylated forms of Pol II, we do think, however, that such analyses go beyond the scope of the present manuscript. Concerning the second remark, we performed H3K36me3 ChIP-on-chip experiments to complement the Pol II inhibition experiments, as suggested by the Referee. As expected, after KM05283 treatment of P5424 cell line we found a significant reduction of H3K36me3 throughout the gene

body of expressed genes (**Fig. 5A and S10D; p. 13, lines 25-27**), suggesting efficient inhibition of transcription elongation.

*5. There are several Figures where controls for tissue specificity would be useful. In Fig. 3B, the scales for H3K4me1 and me3 data are very different. How do these data look when compared with a non-thymocyte (or non-lymphocyte)? Similarly, the modification patterns on Cd3d look quite similar in B versus T lineage cells. Non-lymphocytes apparently were not examined.*

**Response 5:** To comply with these requests, we have now performed a new set of ChIP experiment using chromatin from ES cells as control for tissue specificity along with qPCR analyses of the *Tcrb* gene enhancer (new **Fig. 3B; p. 10, lines 12-13**). As shown in the new **Fig. 3B**, the *Tcrb* enhancer is significantly enriched in H3K4me3 in DRag thymocytes as compared with ES cells (see also **Response 1** to Referee #1 and **Response 6** below). On the other hands, we were surprised by the remark concerning the profiles of the *Cd3d* gene, as this gene displays very distinct profile in B and T cells (compare **Fig. S5B to Fig. 1A**). As described in the result section (**p. 9, lines 26-27**) this example shows that the enhancer of the *CD3d* gene, which is not active in B cells, is associated with H3K4me1/me2, but not with H3K4me3, in this lineage.

*6. A subset of data should be supported by focused ChIP-PCR assays. Most relevant are those shown in Fig. 4, where differences are especially subtle. Are these differences reproducible when performed in a more quantitative manner and how do these regions compare with marks in non-lymphoid cells?*

**Response 6:** As suggested by the referee we have analyzed by ChIP-qPCR the examples shown in **Fig. 4** (new Supplementary **Fig. S8**), along with the known enhancers shown in **Fig. 2** (new Supplementary **Fig. S3**), and compared then with the enrichment found in ES cells or in a gene desert region (NR). This new set of data is described in the result section (**p. 9, lines 7-8 & p. 11, line 24**). Overall, differences in enrichment observed by ChIP-on-chip were validated by ChIP-qPCR. Importantly, the H3K4me3 enrichment observed at known enhancers or tested intergenic regions in DRag and/or DRagCD3 thymocytes were significantly higher than the signal observed in ES cells.

### Referee #3

*In this study, Pekowska et al. nicely document a correlation between histone H3K4 dimethylation and trimethylation and activity of stage-specific enhancers. Figure 1B is particularly revealing, by showing that the previously proposed selective enrichment of H3K4me1 at enhancers is lost when examining stage-specific enhancers. This figure was derived from a genome-wide compilation of modification levels, but the authors then validate this finding by examining individual loci using high-density chip-on-chip tiling arrays. The data provide strong evidence that the authors conclusions are valid at at least a subset of enhancers. The authors also provide a possible connection between H3K4me3 modification of enhancers and polymerase association.*

*This manuscript is highly interesting and will represent an important addition to the literature. It is particularly important because previously published data have led to the erroneous view that all enhancers contain high levels of H3K4me1, but low levels of H3K4me3. However, as with most genome-wide studies, the published studies described statistical averages and trends. By separating stage-specific enhancers from enhancers that are more constitutively active, the authors of the current manuscript show that the H3K4me1/H3K4me3 ratio is much lower at active enhancers. Since many labs are currently identifying enhancers on the basis of a high ratio, this paper will of considerable value.*

*My only suggestion is for the authors to discuss more extensively the observation that H3K4me3 appears to extend through a surprisingly broad region of the active loci. Has this been observed by others? Doesn't the broad distribution raise questions about the hypothesized close link between H3K4me3 and polymerase recruitment?*

**Response 1:** We were pleased with the very positive comments provided by this referee. Concerning the observation highlighted by the referee, this was already mentioned in the discussion

section and has been described in our previous publication (Genome Res. (2010) 20, 1493-502; see also **Response 2** to Referee #1). We have now mentioned this observation in the results section (**p. 6, lines 17-19**) and extended the discussion to address the possible link between the broad distribution of methylated H3K4 and Pol II recruitment and transcription (**p. 18, lines 1-15**).

2nd Editorial Decision

20 May 2011

Thank you for submitting your revised manuscript for consideration by The EMBO Journal. It has been now been evaluated by two of the original referees and I enclose their reports below. As you will see the referees find that the manuscript has significantly improved. However, in contrast to your conclusion referee #1 finds that the ChIP-seq data do not corroborate the broad distributions seen in the ChIP-chip data and questions the resolution of the latter dataset. In order to avoid the potential future criticisms and to cement the importance of the findings in this paper the referee asks that ChIP-seq be performed for H3K4me1 and H3K4me3 in the DN cells. I realize that this is a significant amount of additional work but after discussing this together with Chief Editor, we believe that these criticisms are central to the main finds of the paper and it is in everyone's interest to resolve these issues so the paper can make its important mark in the field.

When you submit a revised version to the EMBO Journal, please make sure you upload a letter of response to the referees' comments. Please note that when preparing your letter of response to the referees' comments that this will form part of the Review Process File, and will therefore be available online to the community. For more details on our Transparent Editorial Process initiative, please visit our website: <http://www.nature.com/emboj/about/process.html>

Thank you for the opportunity to consider your work for publication. I look forward to your revision.

Yours sincerely,

Editor  
The EMBO Journal

-----  
REFeree COMMENTS

Referee #1

The revised version of this manuscript addressed several problems identified in the original paper. Overall, the main conclusion of the manuscript (that active enhancers are often associated with H3K4 di- and tri-methylation) is much more convincing in the current than in the previous manuscript, mainly because of the additional ChIP-Seq data. However, in the current form the manuscript will unavoidably be exposed to strong criticisms: as discussed below, a substantial change in the way data are presented and discussed is required.

The main point relates to the reliability of the ChIP-chip data and the discrepancies with the ChIP-Seq data (now shown only in the supplementary material), specifically regarding H3K4me3 (which is the central part of the story and as such most relevant to the model). H3K4me3 is usually restricted to a few nucleosomes in the area surrounding the transcription start sites of active genes (they may be many nucleosomes in the case of highly active genes, but the overall scenario is the same). When looking at the ChIP-Seq data, this scenario is confirmed. For instance, at the Cd4 gene (Fig S4) there is a nice promoter peak and a smaller, distinct peak overlapping with the Ep enhancer some kilobases upstream. However, the picture reported in the main body of the paper (Fig 2) and obtained by ChIP-chip is completely different, with a very long H3K4me3 streak that (with the exception of interruptions due to the array design at repeats) covers the region upstream as well as the entire gene! Same holds true for Dnnt, which displays a big promoter peak and an upstream, distinct H3K4me3 region of smaller intensity in the ChIP-Seq (Fig.

S4). However, in Fig 1 (obtained by Chip-chip) there is an extended H3K4me3 signal that covers both the upstream region and the 5' half of the Dnnt gene. Moreover, also in this case the difference in intensity between promoter and upstream enhancer cannot be appreciated in the ChIP-chip data. There are also some cases (like CD81) in which the H3K4me3 peaks associated with enhancers have similar height as those associated with promoters, but also here while ChIP-Seq data allow discriminating individual peaks, the ChIP-chip data show a single and extended region of H3K4me3 that makes little sense.

Clearly, the ChIP-chip data in the paper provide a blurred, low-resolution signal and the relative intensities of peaks in ChIP-chip experiments do not reflect what shown by ChIP-Seq, which is way more reliable and quantitative.

However, it would be a pity to give up at this point and my suggestion is that the authors generate two additional data sets (K4me1 and K4me3) in double negative cells and use these data sets (together with the ChIP-Seq data already available) rather than the ChIP-chip data in the main body of the paper.

ChIP-chip data may be used in the supplementary material for ancillary analyses.

Referee #2

The authors have adequately addressed my concerns about the original submission. One minor comment:

It is very difficult to see how the qPCR data shown in Fig. 3B (substantial differences in the genotypes) correspond to the ChIP-chip data in Fig. 3C. (insignificant differences). Either delete Fig. 3C or clarify why these two methods give seemingly different results (perhaps only because the qPCR amplicon is focused to an area of difference).

2nd Revision - authors' response

28 June 2011

### **Point-by-point responses to referees**

**Referee #1**

*The revised version of this manuscript addressed several problems identified in the original paper. Overall, the main conclusion of the manuscript (that active enhancers are often associated with H3K4 di- and tri-methylation) is much more convincing in the current than in the previous manuscript, mainly because of the additional ChIP-Seq data. However, in the current form the manuscript will unavoidably be exposed to strong criticisms: as discussed below, a substantial change in the way data are presented and discussed is required.*

*The main point relates to the reliability of the ChIP-chip data and the discrepancies with the ChIP-Seq data (now shown only in the supplementary material), specifically regarding H3K4me3 (which is the central part of the story and as such most relevant to the model).*

*H3K4me3 is usually restricted to a few nucleosomes in the area surrounding the transcription start sites of active genes (they may be many nucleosomes in the case of highly active genes, but the overall scenario is the same). When looking at the ChIP-Seq data, this scenario is confirmed. For instance, at the Cd4 gene (Fig S4) there is a nice promoter peak and a smaller, distinct peak overlapping with the Ep enhancer some kilobases upstream. However, the picture reported in the main body of the paper (Fig 2) and obtained by ChIP-chip is completely different, with a very long H3K4me3 streak that (with the exception of interruptions due to the array design at repeats) covers the region upstream as well as the entire gene! Same holds true for Dnnt, which displays a big promoter peak and an upstream, distinct H3K4me3 region of smaller intensity in the ChIP-Seq (Fig. S4). However, in Fig 1 (obtained by Chip-chip) there is an extended H3K4me3 signal that covers*

*both the upstream region and the 5' half of the Dnnt gene. Moreover, also in this case the difference in intensity between promoter and upstream enhancer cannot be appreciated in the ChIP-chip data. There are also some cases (like CD81) in which the H3K4me3 peaks associated with enhancers have similar height as those associated with promoters, but also here while ChIP-Seq data allow discriminating individual peaks, the ChIP-chip data show a single and extended region of H3K4me3 that makes little sense.*

*Clearly, the ChIP-chip data in the paper provide a blurred, low-resolution signal and the relative intensities of peaks in ChIP-chip experiments do not reflect what shown by ChIP-Seq, which is way more reliable and quantitative.*

*However, it would be a pity to give up at this point and my suggestion is that the authors generate two additional data sets (K4me1 and K4me3) in double negative cells and use these data sets (together with the ChIP-Seq data already available) rather than the ChIP-chip data in the main body of the paper.*

*ChIP-chip data may be used in the supplementary material for ancillary analyses.*

**Response:** We thank the referee for his/her positive comments about our revised version of the manuscript and additional constructive remarks. We agree with Referee #1 that ChIP-Seq provides more accurate and quantitative data than those obtained by ChIP-on-chip approaches. As suggested by the referee, we have now performed native ChIP-Seq for H3K4me1 and H3K4me3 in DN (DRag) thymocytes and used the complete ChIP-Seq data sets in the main body of the paper. We would like to emphasize that, even though this resulted in substantial changes in the writing and figure setting of the manuscript (described below), all our previous conclusions were confirmed with the new data sets. Figure changes included:

- **Figures 1, 2 and 3** were completely replaced by the corresponding ChIP-Seq profiles and analyses. As also suggested by this reviewer, the original ChIP-on-chip Figures are now shown as supplementary **Figures S1, S3 and S7**, respectively. In Figure 1, we also added an additional example of a recently identified lymphoid-specific enhancer associated to the *Gata3* locus.
- The new examples provided in **Figures 4C and D** were further validated by qPCR (new **Figure S6**).
- **Figure 5C** was also redone using peak selection from the ChIP-Seq data.

To be in line with the new sets of figures and analyses, we have modified several parts of the main manuscript (all highlighted in blue). Notably, these modifications included, mainly:

- Title and Abstract of the manuscript:  
We have modified both the title and the abstract in order to focus on H3K4me3, which now constitutes the central part of the manuscript.
- In the Results section:
  - P. 6, lines 3-4; 7-10; 19
  - P. 7, lines 24-28
  - P. 8, line 5
  - P. 9, lines 7-8
  - P. 10, line 26
  - P. 11, lines 7-23
- In the Discussion section:
  - P. 15, lines 1; 4; 8
  - P. 16, lines 7; 9; 14
- In the Materials and Methods section:
  - P. 19, lines 9-11, 19
  - P. 20, lines 10

Overall, we hope the reviewer will be satisfied with all the changes and new data/figures provided in the current revised manuscript.

## Referee #2

*The authors have adequately addressed my concerns about the original submission. One minor comment:*

*It is very difficult to see how the qPCR data shown in Fig. 3B (substantial differences in the genotypes) correspond to the ChIP-chip data in Fig. 3C. (insignificant differences). Either delete Fig. 3C or clarify why these two methods give seemingly different results (perhaps only because the qPCR amplicon is focused to an area of difference).*

**Response:** We were pleased to read that we have adequately addressed the reviewer's previous concerns. However, we were surprised by his/her new minor comment, which may have come from a slight confusion in our explanations. Indeed, even though the two methodologies used to generate the data shown in **figures 3B and C** have analyzed different areas of the *Tcrb* locus [either "within" (qPCR) or "immediately around" (ChIP-on-chip) the Eb enhancer], they provided consistent results as both supported a local decrease of H3K4me3 at the Eb169 mutant alleles. As a matter of fact, to avoid any confusion and prevent bias due to inherent differences between the wild-type and mutated Eb sequences, the signals originating from the probes that overlapped the core Eb region were ignored in the ChIP-on-chip analyses. This particular point is specified in the Figure Legend (**P. lines 6-9**). We stress the fact that, in both analyses, we did observe a significant difference between the wild-type and mutated alleles. Therefore, we consider that the two methods are consistent and, altogether, strengthen our conclusion. To make the understanding of these experiments more straightforward, we have introduced the following changes on the new **Figure 3C**: (1) the most significant differences between the two samples are now highlighted; (2) the Eb core region is now specified by a grey rectangle; and (3) we have added two related comments in the text of the Results section (**P.10; lines 16-18 and 22**). We hope that this point is now clarified and that the referee will be satisfied with the new changes.

3rd Editorial Decision

04 July 2011

You revised manuscript has been reviewed once more by one of the original referees. On behalf of myself and the referee I would like to thank you for agreeing to extend and strengthen the study by adding the additional ChIP-seq data. As you will see from the comments the referee recommends publication in The EMBO Journal pending some minor text changes, once these are incorporated I will be happy to accept the manuscript for publication.

When preparing your letter of response to the referees' comments, please bear in mind that this will form part of the Review Process File, and will therefore be available online to the community. For more details on our Transparent Editorial Process, please visit our website: <http://www.nature.com/emboj/about/process.html>

Thank you for the opportunity to consider your work for publication. I look forward to your final revision.

Yours sincerely,

Editor  
The EMBO Journal

-----  
REFeree COMMENTS

## Referee #1

I am pleased to see that the authors made a very good job at integrating additional data in the manuscript, whose overall quality is now much higher than the original submission. ChIP-Seq data, in particular, provide a crisper view of chromatin marks both in terms of distribution along the genome and relative intensity, thus making the main conclusions stronger.

I have only one residual point that I think should be more carefully presented, both in the title and in the main body of the manuscript. The concept that H3K4me3 helps identify enhancers that are active at a given point of differentiation is now well substantiated. However, the fact that many active enhancers are marked by H3K4me3 does not imply that all active enhancers bear this mark (and RNA Polymerase II) and therefore that its absence provides evidence of inactivity. In the absence of a direct evidence (which is obviously lacking at this point) that enhancers devoid of H3K4me3 are inactive, the implication of H3K4me3 positivity at enhancers should be discussed more carefully, leaving open the (more than likely) possibility that at least a fraction of H3K4me1 positive and H3K4me3-negative enhancers may be active (see title, abstract, heading of paragraph at page 7, conclusion of the same paragraph at page 8 etc.)

3rd Revision - authors' response

08 July 2011

### **Point-by-point responses to referees**

## Referee #1

*I am pleased to see that the authors made a very good job at integrating additional data in the manuscript, whose overall quality is now much higher than the original submission. ChIP-Seq data, in particular, provide a crisper view of chromatin marks both in terms of distribution along the genome and relative intensity, thus making the main conclusions stronger.*

*I have only one residual point that I think should be more carefully presented, both in the title and in the main body of the manuscript. The concept that H3K4me3 helps identify enhancers that are active at a given point of differentiation is now well substantiated. However, the fact that many active enhancers are marked by H3K4me3 does not imply that all active enhancers bear this mark (and RNA Polymerase II) and therefore that its absence provides evidence of inactivity. In the absence of a direct evidence (which is obviously lacking at this point) that enhancers devoid of H3K4me3 are inactive, the implication of H3K4me3 positivity at enhancers should be discussed more carefully, leaving open the (more than likely) possibility that at least a fraction of H3K4me1 positive and H3K4me3-negative enhancers may be active (see title, abstract, heading of paragraph at page 7, conclusion of the same paragraph at page 8 etc.)*

**Response:** We were pleased to read that we have adequately addressed the referee's previous concerns and we thank the referee for his/her very positive comments about our revised version of the manuscript. Although we agree with the later remark from Referee #1, we would like to point out that, in our previous version of the manuscript, we neither concluded that H3K4me3 is the only mark of active enhancer, nor that the regions enriched in H3K4me1, but devoid of H3K4me3, are necessarily inactive. However, to avoid any remaining ambiguity in our conclusions, we have modified the abstract (P.2, lines 12-14 & 16), several headings of the result section (P.7, line 8; P.10, line 26), the conclusion in page 9 (lines 11-13); and we have added a new paragraph in the discussion section (P.15, lines 5-8) to explicitly discuss this issue. We have not modified the title of the manuscript because we think that, as it is, there is no overstatement ("**H3K4 tri-methylation provides an epigenetic signature of active enhancers**"). Notably, the title does not mean that H3K4me3 is the only signature of active enhancers. We hope that this point is now clarified and that the referee will be satisfied with the new changes.

